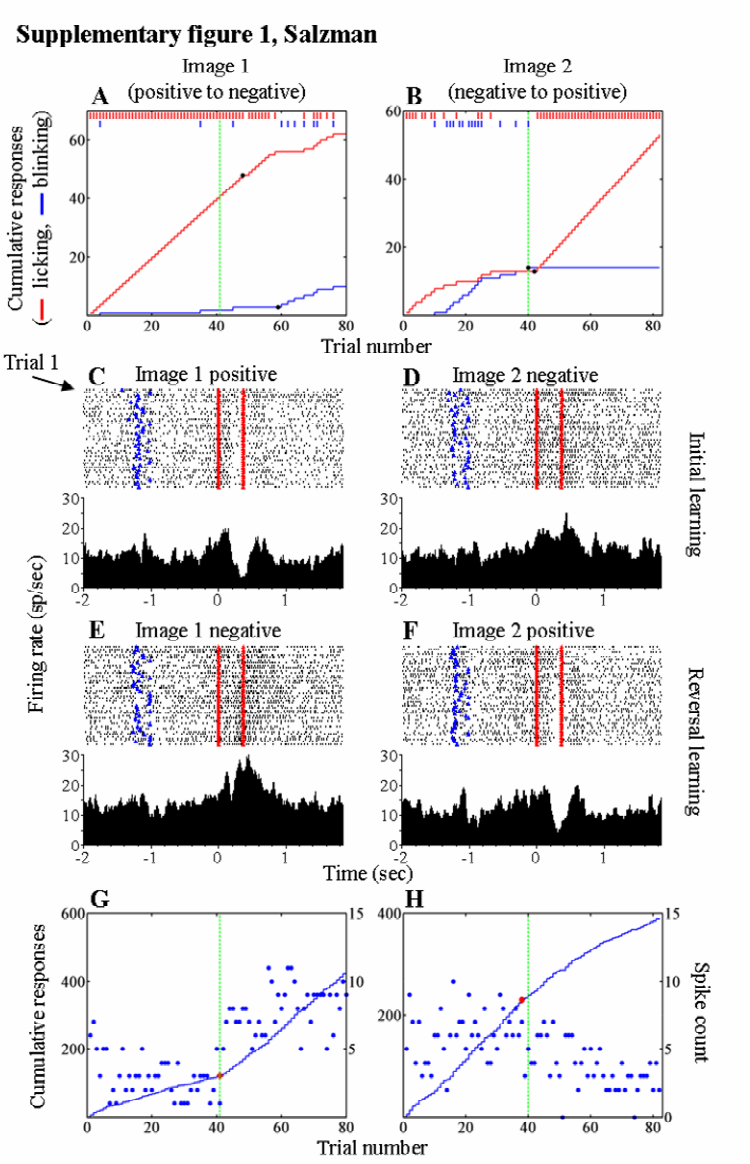
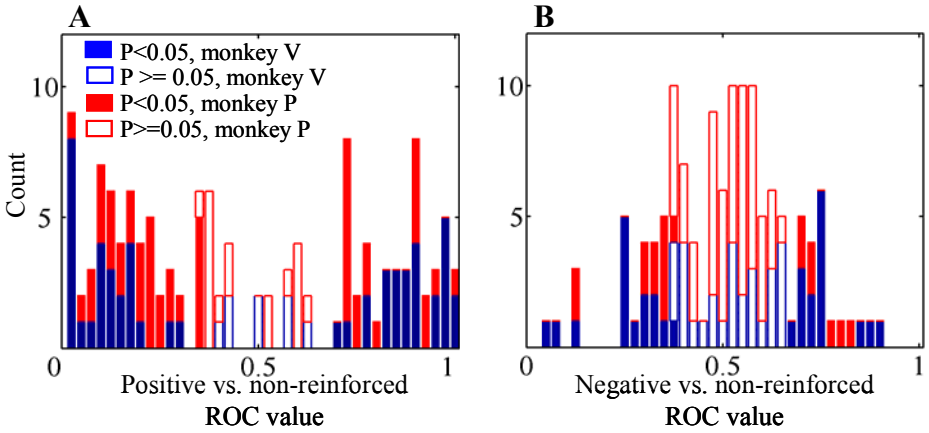


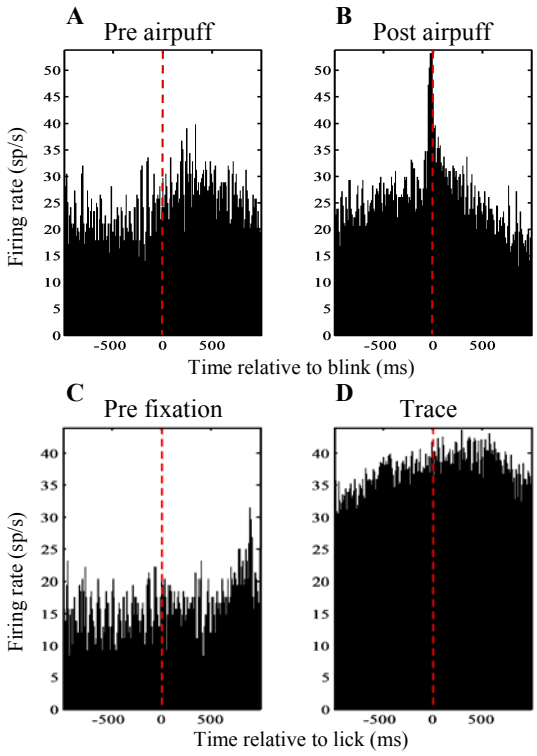
Supplementary Figure 1. Example of an amygdala neuron whose activity reflects value during the visual stimulus interval. This cell responded more strongly when an image was negative than when the same image was positive. **A,B.** Licking and blinking responses (red and blue tick marks, respectively, at the top of the panels), along with cumulative plots of responding, plotted as a function of trial number for both images that underwent a reversal. The black dots indicate the change points for licking and blinking. The change point marks the trial where the slope changes significantly on each cumulative plot. **C-F.** Rasters and PSTHs for the amygdala cell recorded during the learning depicted in **A**,**B**. Left to right tick marks show fixation point onset (blue), and visual stimulus onset/offset (red ticks, aligned in time); the raster is truncated at US delivery. **G,H.** Spike count and cumulative spike count (from the visual stimulus interval) plotted as a function of trial number for Image 1 (**G**) and Image 2 (**H**) plotted continuously across the reversal. Red dots: activity change point, where the slope of the cumulative plots changes. Green vertical line, reversal trial.

Supplementary Figure 2, Salzman

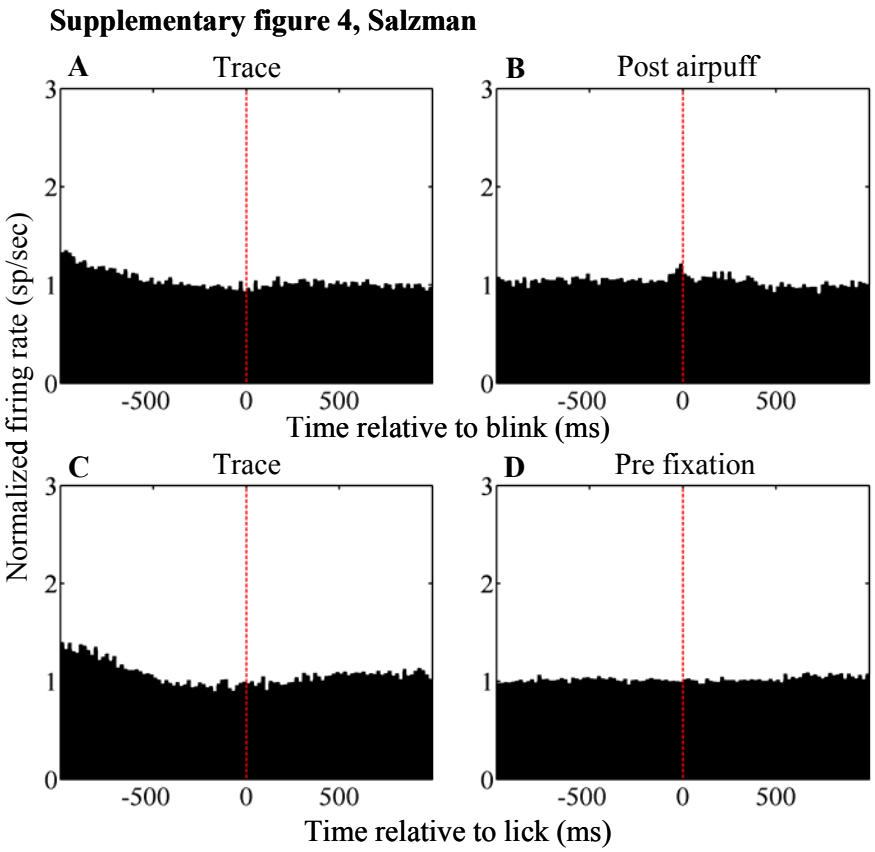


Supplementary Figure 2. Comparison of trace interval activity from non-reinforced images to activity from positive (**A**) and negative (**B**) images. We performed a similar analysis to that shown in Fig. 3C, but which instead compared responses on non-reinforced trials to responses on positive or negative trials in a time period from 350 ms after visual stimulus offset until the time of US delivery (this time period was chosen to eliminate any possibility of a visual response). We selected the cells that responded differentially to positively and negatively conditioned images during the trace interval, excluding cells in which behavior on non-reinforced trials was not different from positive or negative trials (typically, both licking and blinking rates are reduced on non-reinforced trials). We then compared activity from the 20 trials before and after the change in activity related to image value reversal to activity from the non-reinforced image trials interleaved during the same time period. We used the non-reinforced trials as the reference distribution for both comparisons. Each cell contributes up to 2 data points (1 data point per image). **A.** Positive image responses compared to non-reinforced image responses. 104/130 data points had ROC values significantly different from 0.5 (permutation test, $p < 0.05$). **B.** Negative image responses compared to non-reinforced image responses. 47/129 data points have ROC values significantly different from 0.5 (permutation test, $p < 0.05$). We did not perform a similar analysis on visual stimulus activity because those responses are confounded by image selectivity (see Figure 3A).

Supplementary figure 3, Salzman

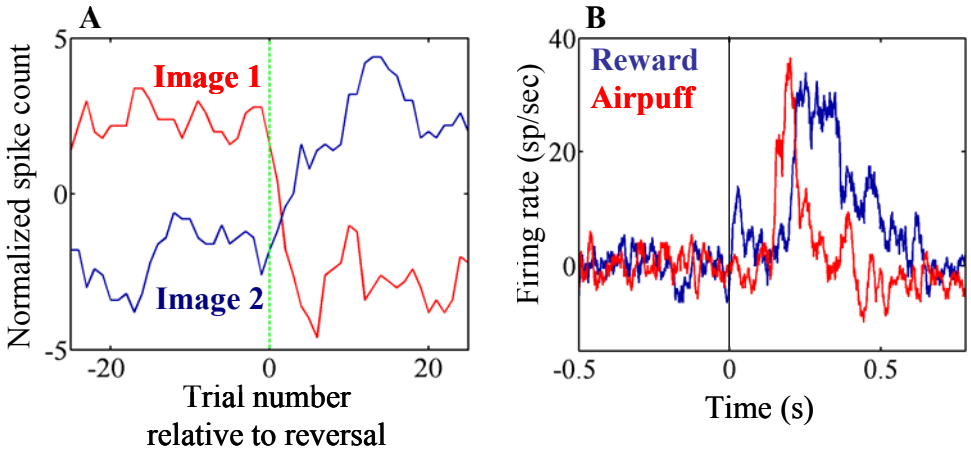


Supplementary Figure 3. Amygdala neural activity is not specifically related to motor responses (licking and blinking) for the cell depicted in Fig. 2. **A,B.** Blink-triggered spike histograms. Blinks were sorted into two categories depending on whether they occurred during the one second before airpuff delivery or the one second following airpuff delivery, and neural activity was aligned on these events. In **B**, the peak in activity just prior to blink onset was a sensory response to air puff delivery, not a response to blinking itself. The peak in activity does not precede blinks that occur during the trace interval (**A**). **C,D.** Lick-triggered spike histograms. Licks were sorted into two categories depending on whether they occurred in the one second before fixation point onset, or the last one second of the trace interval, and neural activity was aligned on these events. In all cases, spikes were binned into 10 ms bins and summed, before smoothing with a 2 bin moving average. The vertical dashed red line marks the time of blink or lick onset, respectively. The cell shows no stereotyped response around the two behaviors, indicating that changes in amygdala responses with learning cannot be attributed to motor signals used for licking and blinking. We performed similar analyses on all the cells in our sample. We used the same latency analysis described in the methods section to look for a response in the 200 ms period preceding a blink or an airpuff compared to a "background" level of activity in the period preceding blinks or licks by 200 – 400 ms. We conducted this analysis separately for blinks and licks occurring during the last 1 sec of the trace interval, and those responses occurring at other times during the trial. We did not find evidence for a consistent motor-related signal in any cell.

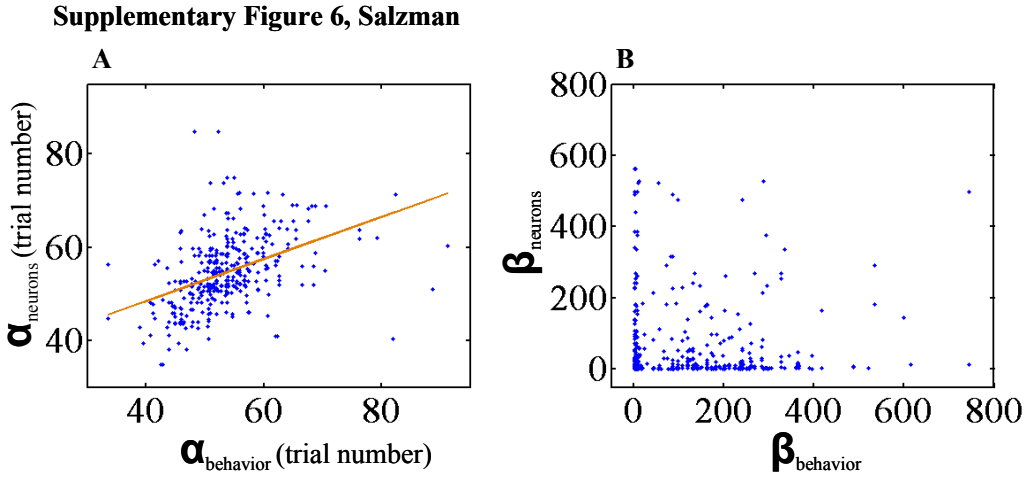


Supplementary Figure 4. Amygdala neural activity is not specifically related to motor responses (licking and blinking) across the population of cells recorded. **A,B.** Blink-triggered spike histograms. Blinks were sorted into two categories depending on whether they occurred during the one second before air puff delivery (trace interval) or the one second following air puff delivery. Neural activity was aligned on these events, normalized by dividing by the median response, and averaged across all neurons. There are no consistent peaks in activity occurring in relation to blinks during these time intervals across the population of neurons. The peak in activity seen before blinks occurring within 1 sec of an air puff is not present in advance of blinks occurring in the 1 sec before air puff delivery. The peak in activity seen for blinks after air puffs is due to a sensory response to the air puff itself that we commonly observe (see Table 1). **C,D.** Lick-triggered spike histograms. Licks were sorted into two categories depending on whether they occurred in the one second before fixation point onset, or the last one second of the trace interval. Neural activity was aligned on these events, normalized by dividing by the median response, and averaged across all neurons. Again, the analysis reveals that there are no consistent peaks in activity in relation to licks for the different time intervals across experiments. For all plots above, spikes were binned into 10 ms bins and summed, before smoothing with a 2 bin moving average. The vertical dashed red line marks the time of blink or lick onset, respectively. Overall, cells show no stereotyped response around the two behaviors, indicating that changes in amygdala responses with learning cannot be attributed to motor signals used for licking and blinking.

Supplementary Figure 5, Salzman

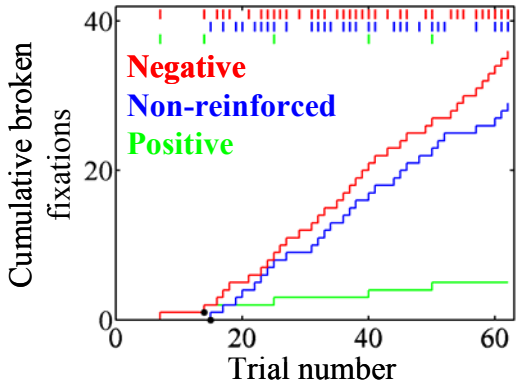


Supplementary Figure 5. Value coding is not simply due to cells receiving information from only reward pathways (for positive cells), or from only pathways representing air puffs (for negative cells). **A.** Normalized neural activity during the visual stimulus interval, smoothed by taking a 5 trial moving average, plotted as a function of trial number for 2 images in one experiment. One image started with a positive value and changed to negative, and the other image started with a negative value and reversed to positive. We normalized activity by subtracting the mean response of each image. This cell responded more when an image was positive than when the same image was negative. The green line demarcates when reversal occurs. **B.** PSTH aligned on the time of reinforcement (time 0) for negative trials (red line; air puff delivered) and positive trials (blue line; reward delivered), with baseline activity from the trace interval subtracted. Although this cell coded the positive value of images, it responded strongly to both air puff and reward.



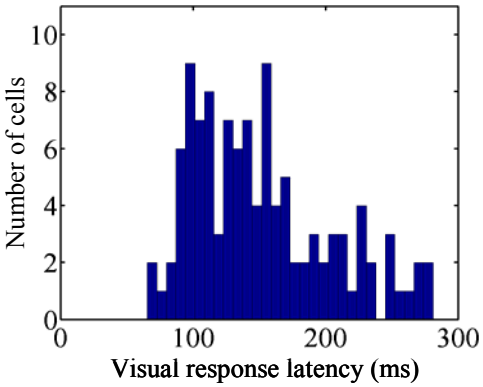
Supplementary Figure 6. The relationship between the trajectories of changes in neural activity and behavioral learning on a cell by cell basis. In addition to fitting sigmoid functions to the average normalized data across our population of value-coding cells, we also fit the same functions to data from the 20 trials before and after reversal for each experiment. Each value-coding neuron contributed up to 4 data points for this analysis (1 for each image from the visual stimulus and trace interval). **A.** α from the fit to behavioral data plotted against α from the fit to neural data. The estimate of α for neurons and behavior was significantly correlated across experiments ($p < 0.0001$, $r^2 = 0.16$), just as the onset of changes in neural activity and behavior were correlated. For this plot, we excluded 69 data points where the estimate of α was greater than 100, which corresponded to cases where either behavioral or neural changes occurred long after reversal, or when data was particularly noisy. **B.** β from the fit for behavior plotted as a function of β from the fit of neural activity. The data show that many β estimates have quite high values, corresponding to steep slopes in the fit sigmoids. We frequently observed this phenomenon when changes in behavior or neural activity appeared to be like a step function.

Supplementary Figure 7, Salzman



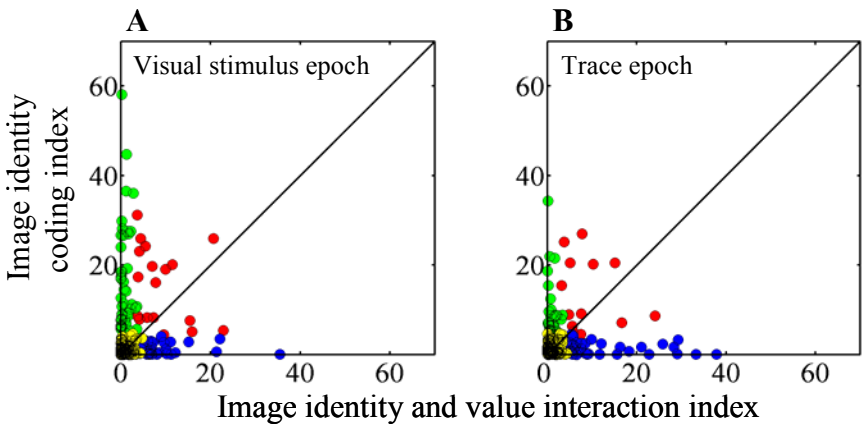
Supplementary Figure 7. Broken fixation behavior increases when a monkey learns the negative or non-reinforced value of a visual stimulus. For this experimental session, we presented the visual stimulus for 1 sec, and broken fixation behavior typically occurred 400 – 750 ms after image onset. Broken fixation and cumulative broken fixations plotted as a function of trial number for positive (green), non-reinforced (blue), and negative (red) image trials. The tick marks indicate trials where the monkey broke fixation. A significant increase in broken fixations occurs after 15 trials for non-reinforced images, and after 14 trials for negative trials ($p < 0.05$, change point test). These data demonstrate that the monkey tries to avoid the non-reinforced and negative trials by breaking fixation.

Supplementary Figure 8, Salzman



Supplementary Figure 8. Histogram of the visual response latencies, computed separately for each cell (N = 196 cells; see Supplementary Methods for details of analysis).

Supplementary Figure 9, Salzman



Supplementary Figure 9. Image Identity Coding Index plotted against Image Value-Identity Interaction Coding Index for the visual stimulus (**A**) and trace (**B**) intervals for the same 2-way ANOVA shown in Fig. 3A,B. Blue dots, $p < 0.05$, image value-identity interaction coding index. Green dots, $p < 0.05$, image identity coding index. Red dots, $p < 0.05$ for both factors. Yellow dots, n.s. The index values correspond to the percentage of variance accounted for by image identity and by the interaction between image identity and value, respectively.

Supplementary Notes

Supplementary Note 1. We refer to changes in activity related to image value reversals as "value coding" because they occur as monkeys learn the association between a CS and a positive or negative US. Since the USs we employ have either a positive reinforcing value (eliciting licking, an approach behavior) or negative reinforcing value (eliciting blinking, a defensive behavior), changes in CS-related activity upon learning an association reflects the fact that the CS has acquired predictive value to a monkey. If a cell changed activity in the same direction for both images upon image value reversal, we did not classify it as a value-coding cell. Such responses could be related to other processes important for learning. For example, activity that changes in the same direction may reflect arousal or attention, which could be elicited by a change in task contingencies.

Supplementary Note 2. We also characterized trace interval responses on non-reinforced trials compared to responses on negative and positive trials (see Supplementary Figure 2). Although many amygdala neurons differentiated between these conditions, in general, responses to non-reinforced images were more similar to responses to negative images than to positive ones. This finding is consistent with the notion that non-reinforced trials are mildly aversive, since monkeys are normally motivated by rewards. Indeed, when visual stimuli are presented for longer intervals, in both our paradigm and in other studies, monkeys try to avoid trials rewarded less[1] or not at all[2] by either breaking fixation or failing to complete trials correctly (see Supplementary Figure 7).

Supplementary Note 3. Although functional neuroimaging in humans has illuminated aspects of how the amygdala processes visual stimuli in relation to value, the limited spatial and temporal resolution of fMRI cannot elucidate whether individual amygdala neurons respond differentially to positive and negative visual CSs, as we have done here.

Supplementary Methods.

General Methods.

During experiments, monkeys sat in a Plexiglas primate chair (Crist Instruments) with their eyes 57 cm in front of a 21" Sony CRT monitor. Monkeys were under constant visual observation by the experimenter by way of an infrared video camera connected to a monitor that displayed the monkey to the experimenter.

Electrophysiological Recording and Experimental Control

We used the TEMPO (Reflective Computing) package of software running on Dell Optiplex 260 PCs for experimental control, and the Plexon system for neurophysiological recording, signal amplification, filtering, digitizing of spike waveforms, and data collection. The TEMPO software program directed the sequence of events in a trial, governed stimulus presentation, and enforced the behavioral demands of the task. Analog signals, such as those representing eye positions, were directed to A/D boards in the TEMPO system for behavioral control, and in the Plexon system for subsequent data storage and analysis. The TEMPO system sent "event codes" in real time to the Plexon system through a digital I/O interface, so that the Plexon data files contained all data from an experiment.

Visual stimulus presentation was accomplished with the Videosync program available with TEMPO that runs on a dedicated slave PC. Through a digital I/O interface, the TEMPO server PC directs the Videosync program as to what stimulus to deliver and when to display it. The Videosync computer sent a TTL pulse to the TEMPO computer at the beginning of a vertical retrace in which the display changed. Visual latency analyses occurred relative to this TTL pulse. During periods of fixation, the monkey was required to maintain a position of gaze within 3.5 degrees of the fixation spot, as measured with an infrared eye tracker (ASL, Applied Science Laboratories) that captured pupil images at 240 Hz. Images typically occupied an 8 degree square centred over the fovea.

We measured licking by placing the liquid reward tube just away (~1-2 cm) from the monkey's mouth. Every ms, we measured whether the monkey's tongue interrupted an infrared beam of light that passed between the monkey's mouth and the lick tube. During blinks, the eye tracker would lose its signal and output a characteristic voltage, which we identified and analyzed in Matlab. We verified that these signals were blinks by comparing the signals to a video camera feed showing the monkey blinking during experiments.

The location of the amygdala was ascertained in stereotactic coordinates using MRI imaging with each monkey anaesthetized with isoflurane, intubated and monitored in a 1.5 Tesla research magnet in the Columbia University Department of Radiology. We used these images to guide our subsequent placement of recording chambers.

After a 1-2 week postoperative recovery, we returned to the MRI scanner, verified that our chamber was placed over the amygdala, and placed an electrode in the brain dorsal to the amygdala. To position the electrode, we advanced it through a guide tube that rested in a Crist grid and transversed dura (Crist Instruments). The Crist grid has holes in a grid formation separated by 1 mm. We placed the electrode a known distance

into the brain and acquired images which visualized the electrode in the brain directed at the amygdala. We used two different sequences to visualize the amygdala and tungsten electrodes. A 2D SPGR sequence, with 0.7mm slice thickness, 0 intergap spacing, and 0.234 mm X 0.234 mm within slice pixels, easily revealed the amygdala and a large susceptibility artefact from an electrode (e.g. Fig. 1B). A 2D IR sequence, with 2mm thick slices and 0 intergap spacing and 0.234 mm x 0.468 mm within-slice resolution, still visualized electrodes (e.g. Fig. 1C), but also revealed some anatomical structure within the amygdala (e.g. Fig. 1D). A fluid-filled Crist grid was visible at the top of images using both sequences (see Figs. 1C,D). From the MRIs we obtained, we calculated the distance that the electrode must travel from the end of the Crist grid and the end of the guide tube to the amygdala, and we also noted grey and white matter transitions to guide electrode positioning. So that we could position the electrode in the brain at coordinate locations as determined in the MRI scanner, we also used a grid that supported guide tubes during recording sessions in the lab. We employed either the Crist grid (when recording with one electrode), or a grid supplied with the NAN microdrive in which grid holes were spaced 1.3 mm apart. 4 electrodes separated by 0.34 mm were independently advanced through a single guide tube in the NAN grid.

Data Analysis

Visual Response Latency. To estimate visual response latency, we compared activity during the 500 ms preceding visual stimulus onset to activity during the presentation of visual stimuli. First, for the activity preceding fixation point onset, we constructed a histogram of the number of spikes in each 20 ms bin, shifted by 1 ms across the entire 500 ms period. We then defined a criterion response as one that exceeded 99% of the bins in the histogram (for excitatory responses) or that was less than 95% of the bins (for inhibitory responses). Next, we determined which 20 ms bins, slid in 1 ms steps, met the criterion response during the time from visual stimulus onset until stimulus offset. We defined latency as occurring at the beginning of the first of 20 consecutive overlapping bins that met a criterion response. We computed the latency separately for each image for all cells we recorded. If a cell responded to more than one image, we took the average latency as the estimate of response latency for that cell. We selected 90 ms after visual stimulus appearance as the beginning of the visual stimulus interval because ~95% of the latencies were greater than 90 ms (Supplementary Figure 8). This approach ensured that we would analyze the exact same time intervals in every experiment.

ANOVA Analysis. To determine the relative contribution of image identity and image value to neuronal responses, we performed a 2-way ANOVA on every cell, analyzing the responses to both images that reversed in value. For each cell, this analysis compared activity from the 20 trials before and after the identified change point for each image. If the change point test detected a significant change in activity for only one image, then for the other image we examined activity from the 20 trials before and after the change point identified for the first image. If no change points were detected for either image, then we compared activity from the 20 trials before and after reversal for both images.

ROC Analysis. To determine whether a given cell responded more strongly when an image had a negative or positive value, we performed an ROC analysis that compared activity from the 20 trials before and after an identified change point for a given image. We always designated the negative value trials as the reference distribution. Consequently, ROC values > 0.5 indicated that a cell responded more when an image was positive than when the same image was negative. By contrast, ROC values < 0.5 corresponded to cells that responded more strongly when an image was negative compared to positive. 5 cells changed activity in opposite directions in the visual stimulus and trace interval (ROC values significantly different from 0.5), and we excluded those cells from all analyses.

Permutation Test on Computed ROC Values. We used a permutation test to evaluate whether each ROC value was significantly different from 0.5 (Fig. 3C). For this analysis, we abolished the relationship between neural activity and image value by randomly assigning each neural response a value, but keeping constant the overall distribution of neural activity and the number of responses assigned to each value. We then computed the ROC value for this permuted data set, and repeated the procedure 1000 times. An ROC value on the original data set that fell outside the central 95% of the distribution of ROCs computed on the permuted data was considered statistically significant.

Normalization Procedure. When comparing activity across the population of neurons, we normalized activity for each cell by first dividing each trial's activity by the median response across trials, and subsequently subtracting the mean of the normalized values across all trials from each trial's normalized value. For all images, we analyzed activity from the 20 trials before and after the reversal in image value. Note that we also multiplied by -1 the normalized activity from the neurons responding more strongly before reversal on both reinforced and non-reinforced trials (so that activity went from low-to-high for all cells). Since all trial types were randomly interleaved, the change in activity on positive and negative images could not be due to changes in overall cell responsivity as a result of the reversal, or we would have observed similar changes on the non-reinforced images. We performed a similar procedure on the behavioural data, except normalizing by the mean of the data values. We fit Weibull functions to the behavioural and neural data, with four free parameters (see equation 1). l and u parameters adjusted the lower and upper asymptotes, and α and β influenced the shape of the function (α adjusted the latency of the rise of the function, and β the abruptness of the onset of the rise). Population learning curves for licking and blinking were extremely similar, so they were combined for this analysis (95% prediction intervals of the two curves overlapped at every trial). Similarly, the 95% prediction intervals for curves describing neural activity and behaviour overlapped at every trial for visual stimulus and trace interval activity both for images that changed from positive-to-negative and for images that changed from negative-to-positive. Consequently, we combined all these data to yield one neural learning curve and one behavioural learning curve (Fig. 4C).

References

1. Roesch, M. R. & Olson, C. R. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304**, 307-10 (2004).
2. Shidara, M., Aigner, T. G. & Richmond, B. J. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *Journal of Neuroscience* **18**, 2613-25 (1998).